14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

SURROGATE MODELING USING GAUSSIAN PROCESS REGRESSION FOR ANALYZING UNCERTAINTY ASSOCIATED WITH VARIOUS TSUNAMI SCENARIOS OF AN ELECTRICAL POWER INFRASTRUCTURE

AYUMI NISHI[1] AND GAKU SHOJI[2]

[1]Tokyo Electric Power Service Co., Ltd.
Postal Address
E-mail ayumi-nishi@tepsco.co.jp

[2]University of Tsukuba
Postal Address
E-mail gshoji@kz.tsukuba.ac.jp

Key words: Tsunami, Probabilistic Tsunami Hazard Assessment, Tsunami Scenarios, Uncertainty, Surrogate Model, Gaussian Process Regression

Abstract. The final objective of this study is to develop tsunami scenarios for coastal power plants that include site-specific uncertainties related to the maximum tsunami height, periodic variations, and the direction of tsunami. Initially, a framework was presented to quantify the uncertainties of the tsunami scenarios. Surrogate models for tsunami numerical calculations were constructed targeting the maximum tsunami height at the coastal line locations along the Pacific coast for the wave sources along the Japan Trench. In constructing the surrogate model, features representing the source characteristics in tsunami numerical calculations and the relative location to tsunami source and evaluation point were quantified, and the relationship between these features and the maximum tsunami heights at the coastal line was modeled using Gaussian process regression. Compared to discrete tsunami numerical calculation results, the surrogate model can provide a more detailed distribution of maximum tsunami height under continuous variation of input parameters. Through a detailed sensitivity analyses of parameters using the surrogate model, dominant parameters for the maximum tsunami height were identified. Furthermore, sensitivity analyses of the tsunami hazard curves generated by the surrogate model were conducted.

1 INTRODUCTION

For a reliable assessment of tsunami disaster at coastal power plants, it is essential to quantitatively evaluate the uncertainties inherent in tsunami scenarios. This is especially true for high-risk facilities such as nuclear power plants, where scenario uncertainties must be thoroughly addressed. Studies such as those by Kotani et al. [1], Gopinanthan et al. [2], Snelling et al. [3], and Salmanidou et al. [4], have demonstrated the uncertainties in tsunami scenarios using surrogate models for tsunami numerical calculations. Kotani et al. [1] proposed a method

for generating hazard curves using response surfaces of tsunami source parameters and tsunami heights, quantifying the contributions of source parameter variation to the response uncertainties. For probabilistic assessments considering uncertainty in inland inundation, Williamson et al. [5] and Fukutani et al. [6] provide relevant research. Fukutani et al. [6] conducted hazard assessments of inland inundation depth distributions, considering the failure probability of seawalls, using an inundation depth distribution generation model by applying proper orthogonal decomposition (POD).

Comprehensive uncertainty evaluation necessitates the systematic classification and probabilistic quantification of uncertainties stemming from tsunami sources characteristic, such as source locations and fault rupture distributions, as well as uncertainties related to coastal and site topography. Based on these uncertainties, it is desirable to set Tsunami scenarios for inundation assessment at the site. Kihara et al. [7] and Nishi et al. [8] demonstrated a method for creating tsunami scenarios for evaluating inundation inside the seawall, based on hazard curves of maximum tsunami heights at the front of seawall at a hypothetical nuclear site. There is a demand for inundation scenario assessment methods that include not only maximum tsunami heights but also the uncertainties of periodic variations and tsunami directions.

The final aim of this study is to construct tsunami scenarios that include site-specific uncertainties of the wave fields related to maximum tsunami heights, periodic variations, and tsunami directions. Chapter 2 proposes a framework for quantifying the uncertainties of tsunami scenarios for coastal power plants. In the following sections (Sections 3 onwards), based on case studies targeting hypothetical earthquakes along the Japan Trench, surrogate models for tsunami numerical calculation were constructed, and an attempt was made to evaluate the uncertainties regarding maximum tsunami height using the surrogate models.

2 A FRAMEWORK FOR QUANTIFYING THE UNCERTAINTY IN TSUNAMI SCENARIOS FOR COASTAL POWER PLANT

A framework for quantifying the uncertainties in tsunami scenarios for coastal power plant is presented in Figure 1. This framework assesses uncertainties at two locations: the shoreline and inside the power plant site. Uncertainties are evaluated separately at each location because the dominant factors to the response are expected to differ due to the complex behavior of tsunamis influenced by the land topography.

In “(A) uncertainty assessment on coastal line”, a surrogate model using tsunami numerical calculations is employed for more detailed and reliable uncertainty quantification. First, tsunami numerical calculations are used to figure the response waveforms at evaluation points on the coastal line. Next, features contributing to response uncertainty are extracted from the input data of the tsunami numerical calculation. Specifically, this includes features representing the tsunami source characteristics (fault parameters), the positional relationship between the source and evaluation points, and the coastal topography. A Gaussian process regression surrogate model is then constructed, using these extracted features as input variables and the response characteristics at the coastal line as the target variable. The response characteristics considered include maximum tsunami height, periodic variation, and tsunami direction. By inputting numerous randomly generated samples of input variables into the constructed surrogate model, continuous and detailed probability distributions can be generated, compared to the probability distribution obtained directly from the tsunami numerical simulations. For

example, in the studies from Chapter 3 onwards, a surrogate model constructed using 990 cases of tsunami numerical calculations for the tsunami source JTT (Figure 2 (ii)) was used to obtain a probability distribution of maximum tsunami height for 10,000 cases. Next, sensitivity analyses of the features is performed using the constructed surrogate model to calculate the contribution of each feature to the variability of the response characteristics. Finally, based on the sensitivity analysis results, sensitivity analysis for the probabilistic tsunami hazard curve is performed using the surrogate model.

In "(B) uncertainty assessment on site", uncertainty in response characteristics due to site-specific factors (e.g., seawall shape) is evaluated. By comparing the uncertainty assessments on the coastal line and inside the power plant site, the uncertainties caused by the tsunami source, coastal topography, and site topography can be classified and quantified, and incorporated into the tsunami scenario assessment.

The following studies focuses on the uncertainty quantification of the maximum tsunami height at coastal line as part of the overall framework. Specifically, the relationship between source characteristics and maximum tsunami height is modeled using tsunami numerical simulation results as training data. Sensitivity analysis of the maximum tsunami height and hazard curve is then performed using the constructed surrogate model.

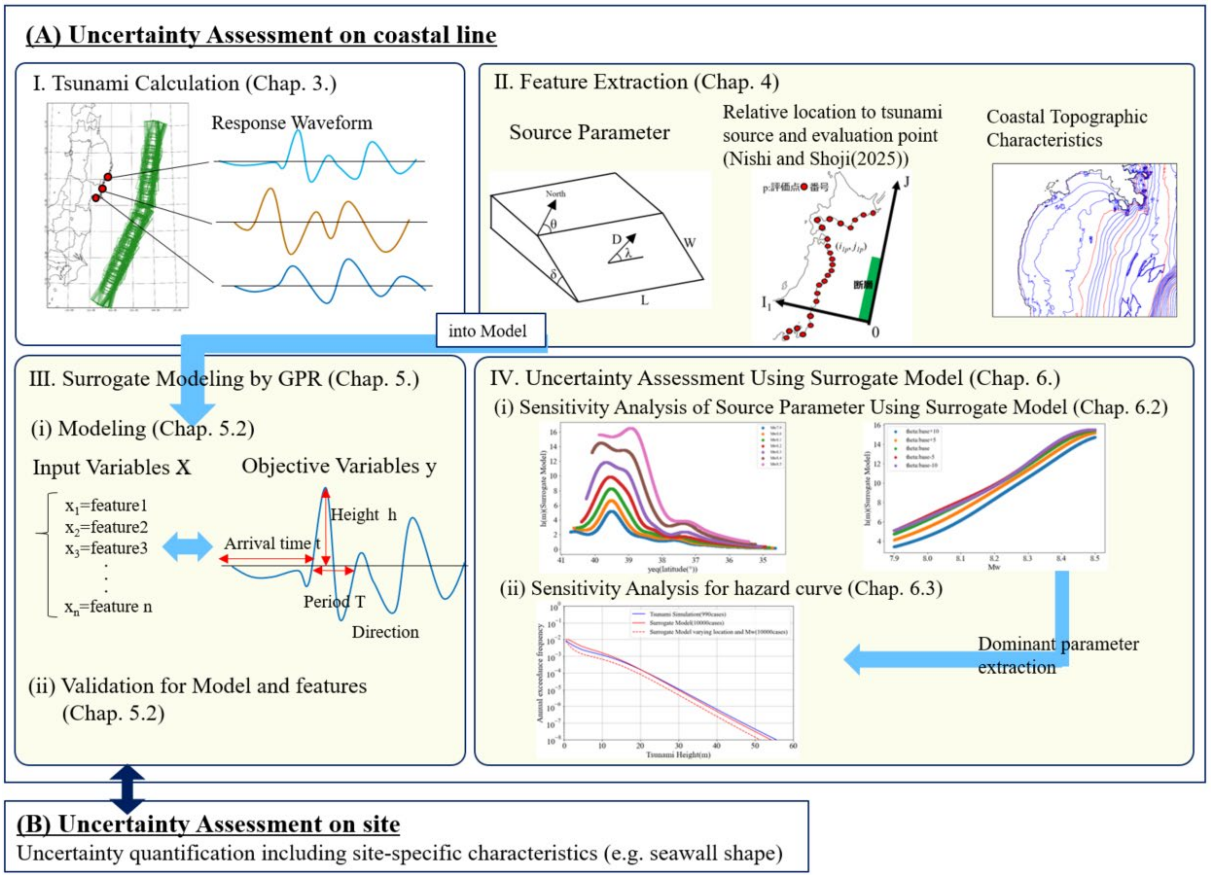
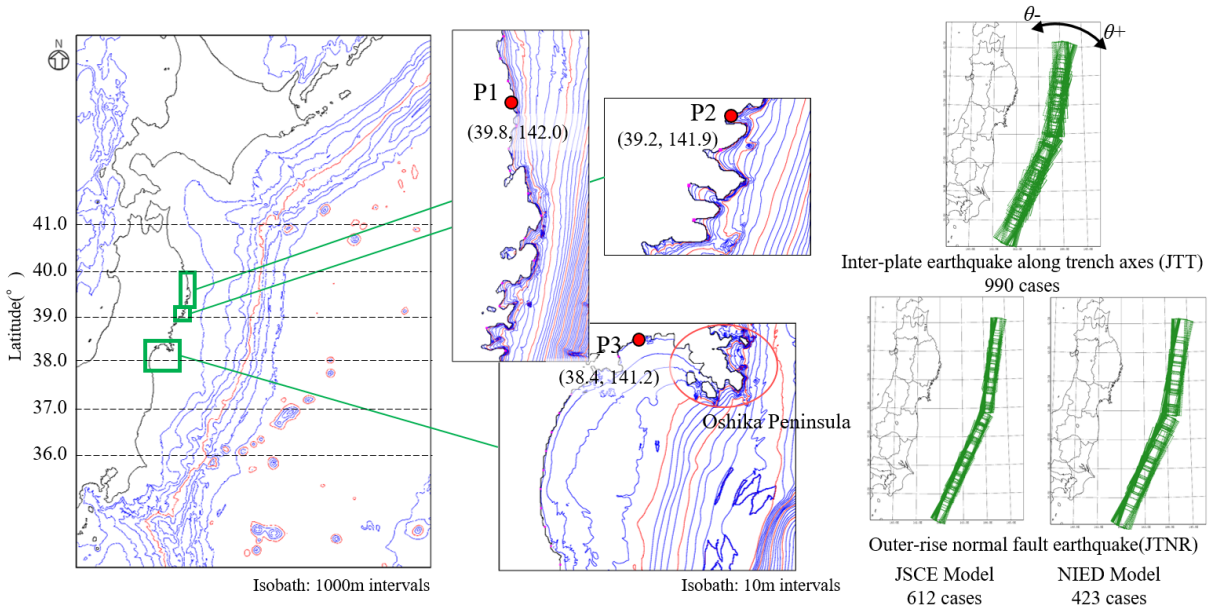


Figure 1: A framework for quantifying the uncertainties in tsunami scenarios for coastal power plant



(i) Evaluation Points (latitude, longitude) (ii) Sources

Figure 2: Evaluation points and Sources

3 A DATASET GENERATED THROUGH TSUNAMI NUMERICAL CALCULATION

Our study conducts tsunami numerical calculation using nonlinear long-wave equations wave equation to investigate tsunamis generated along the Japan Trench. The tsunami sources considered are interplate earthquakes occurring near the Japan Trench (hereinafter referred to as JTT) and outer-rise earthquakes along the Japan Trench (hereinafter referred to as JTNR) (Figure 2 (ii)). The reference fault models are based on previous tsunami reproduction models from JSCE [9]. For JTT, the 1896 Meiji Sanriku earthquake tsunami fault model was used. For JTNR, the 1933 Showa Sanriku earthquake tsunami fault model was used. In addition to the above, JTNR model referenced from the Japan Trench outer-rise earthquake model by NIED [10] was also used as a reference fault model. The uncertainties in the source models, as shown in Table 1, include location, M_w, strike angle θ, and fault depth d for JTT, and location, Mw, strike angle θ, and upper-edge depth d, and the reference fault model for JTNR. Regarding fault scaling, the fault width W was fixed at 50 km, and the fault length S and slip D were adjusted to maintain a constant ratio of S: D according to M_w. For the fault plane angle, the dip angle δ was fixed at 20° for JTT and 45° for JTNR, and the slip angle λ was set according to the strike θ. Note that the basic strike differs between the northern and southern parts of the Japan Trench, as shown in Figure 2 (ii). When the fault spanned both the northern and southern sections of the Japan Trench, two fault models were implemented.

To compare the impact of topographic differences on tsunami response uncertainty in each area, multiple evaluation points were configured. As shown in Figure 2 (i), three points were selected on the coastal line facing the Pacific Ocean. P1 is located in Iwate Prefecture. P2 is located in a bay with a complex coastline in the Sanriku coastal area of Iwate Prefecture, where the topographic amplification effect on tsunamis is significant. P3 is located in the Sendai Plain of Miyagi Prefecture. At P1 and P2, tsunamis propagating directly from the source arrive from

approximately the east. However, at P3, the tsunamis arriving from the east are diffracted around Oshika Peninsula before reaching the location. Tsunami numerical calculations based on nonlinear long-wave equations were conducted for each source, and the maximum tsunami height at each evaluation point was extracted. The boundary conditions were set to ensure complete reflection of the tsunami at the land boundary.

Table 1: Uncertainty of fault model (for tsunami numerical calculation)

Source	Uncertainty				
	Source location	M_w	Strike θ	Fault depth d	Model
JTT	Along JapanTrench	7.9-8.5 (0.1 intervals)	base, base±5, base±10(°)	0, 1.0, 2.0 (km)	JSCE(2016)
JTNR	Along JapanTrench	8.0-8.6 (0.1 intervals)	base, base±5(°)	0, 1.0, 2.0 (km)	JSCE(2016) or NIED(2015)

Table 2: Features (input valiables for surrogate models)

①Source characteristic	②Positional relationship between the source and evaluation points
Edge point of the fault (x, y), Center point of fault (x_{eq}, y_{eq}), Fault length L, Fault width W, Strike θ, Fault depth d, Dip angle δ, Slip angle λ, Slip D, M_w	Relative location to tsunami source and evaluation point P $i_{1p}, j_{1p}, i_{2p}, j_{2p}$

4 IDENTIFICATION OF INPUT FEATURES FOR THE SURROGATE MODEL

Table 2 lists the features considered in this study. Features representing source characteristics include the edge point of the fault (x, y), center point of fault (x_{eq}, y_{eq}), fault length L, fault width W, strike angle θ, fault depth d, dip angle δ, slip angle λ, slip D, and M_w. Features representing the positional relationship between the source and evaluation points were calculated using "the relative location to tsunami source and evaluation point", $i_{1p}, i_{2p}, j_{1p}, j_{2p}$ proposed by Nishi and Shoji (2025) [11]. Nishi and Shoji (2025) [11] demonstrated the validity of these features by showing that incorporating them into a support vector machine [12]-based tsunami height prediction model improved its accuracy.

5 SURROGATE MODELING BY GAUSSIAN PROCESS REGRESSION AND ITS ACCURACY EVALUATION

5.1 Gaussian process regression

Gaussian process regression (GPR) is a non-parametric regression method that can flexibly model complex relationships between inputs and responses without defining a functional form for the regression surface [13]. Assuming that the response $\boldsymbol{y}$ corresponding to the input $\boldsymbol{x}$ follows a Gaussian distribution, the predictive distribution of $\boldsymbol{y}$ is defined as follows:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right) \tag{1}$$

Here, $\boldsymbol{\mu(x)}$ is the mean of input $\boldsymbol{x}$, and $k(\boldsymbol{x}, \boldsymbol{x}')$ is the kernel function. In this study, the Matérn

5/2 kernel was employed for $k(\boldsymbol{x}, \boldsymbol{x'})$.

$$k(\boldsymbol{x}, \boldsymbol{x}') = {\sigma_f}^2 \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) exp\left(-\sqrt{5}r\right) \tag{2}$$

Here, σ_f is a hyperparameter. r represents the distance between the m-dimensional input variables x_m, defined in Equation (3).

$$r = \sqrt{\sum_{m=1}^{M} \frac{(x_m - {x_m}')^2}{l_m}} \tag{3}$$

The hyperparameter l_m represents the length scale. l_m is defined for each dimension m of the input variables and represents the weight for each input variable in the model. For model implementation in this study, we used the Gaussian process regression library GPy [14] in Python, and the L-BFGS-B method implemented in Gpy was used as the optimization algorithm.

5.2 The cases and accuracy evaluations

For each evaluation point, a model was developed to represent the relationship between the maximum tsunami height and the features (Table 1). A separate model was built for each evaluation point and each source zone. Given three evaluation points and two source zones (JTT and JTNR), a total of six models were constructed. In this study, 5-fold cross-validation was performed using K-fold method to demonstrate the reliability of the models. The root mean square error (RMSE) was adopted as the metric for predictive accuracy. Table 3 shows the RMSE for the models at each evaluation point. All models achieved RMSE values of approximately 0.1 m or less, validating their use.

Table 3: Prediction accuracy of the models at each evaluation point

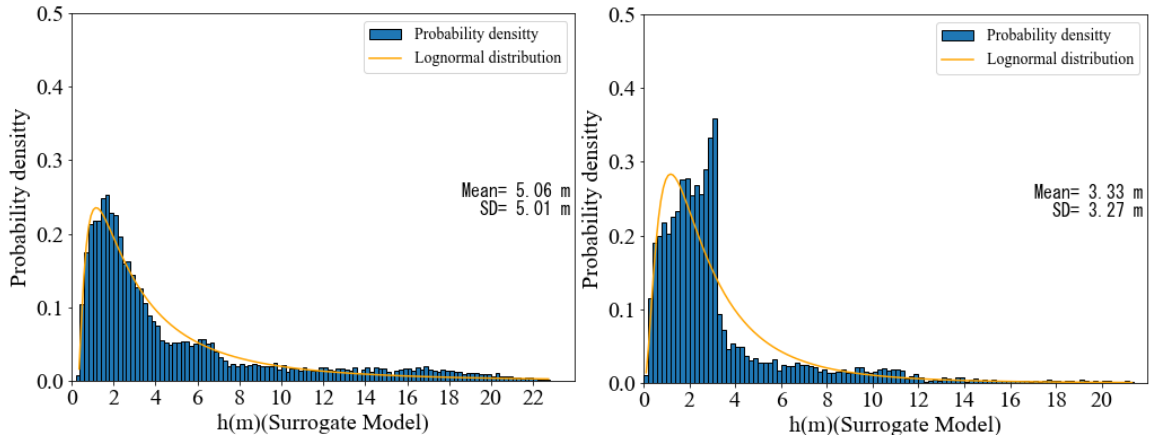
Evaluation point *P*	*RMSE*	
	JTT	JTNR
P1	0.056	0.140
P2	0.032	0.068
P3	0.105	0.071

6 UNCERTAINTY ASSESSMENT USING SURROGATE MODELS

6.1 Probability distributions of maximum tsunami height using the surrogate models

Using the surrogate models developed in Chapter 5, the probability density function (PDF) of the maximum tsunami height was generated. Specifically, a dataset of 10,000 randomly sampled input variable sets was created within the uncertainty ranges of the input parameters listed in Table 1 (source location, M_w, θ, d, model), and these sets were input into the surrogate model. Where there were correlations among some of the input variables, the values of these correlated variables were varied accordingly. For example, the edge point of the fault (x, y) and the relative location to tsunami source and evaluation point i_{1p}, i_{2p}, j_{1p}, j_{2p} were set according to the center point of fault (x_{eq}, y_{eq}) of the source. While the tsunami numerical calculation

parameters used for training the surrogate model were discrete (e.g., M_w discretized in 0.1 increments within the range of 7.9-8.5), the input parameters for the surrogate model were generated continuously. This allowed for the generation of a more detailed probability distribution of the maximum tsunami height. Figure 3 shows the PDF of the maximum tsunami height at P2 generated using the surrogate model. For JTT, the mean tsunami height was 5.06 m with a standard deviation of 5.01 m; for JTNR, the mean was 3.33 m with a standard deviation of 3.27 m. The higher tsunami heights for JTT are attributed to the larger slip amounts in JTT compared to JTNR for fault models with the same M_w.

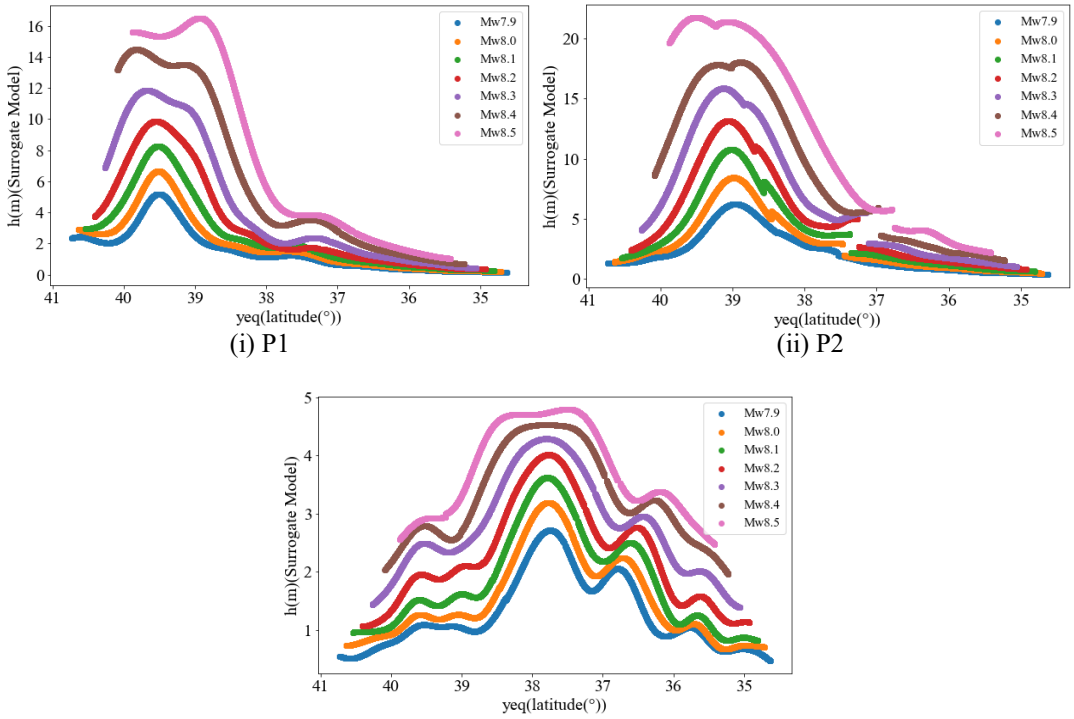


(i)JTT (ii)JTNR

Figure 3: Probability density of tsunami height h at P2 by surrogate model (10000 cases)

6.2 Sensitivity analyses for features

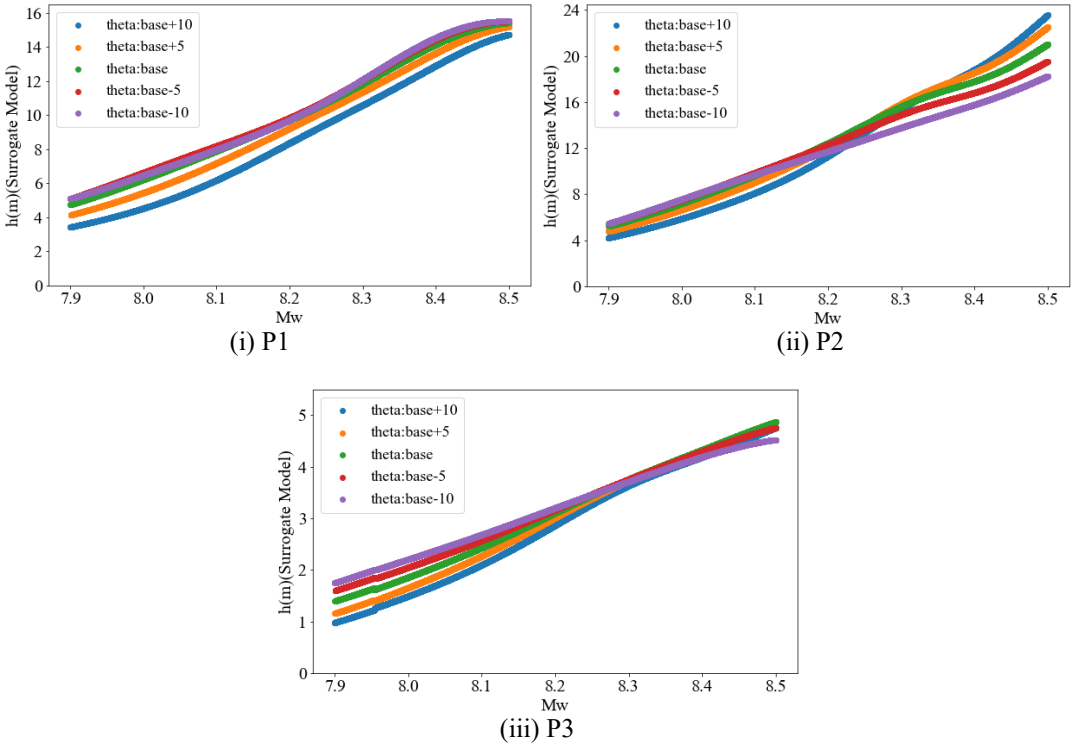
Sensitivity analyses for tsunami height with respect to the uncertainties presented in Table 1 were performed using the surrogate models constructed in Chapter 5. Following the same methodology as in Chapter 6.1, datasets were created by varying only the parameters under consideration for each sensitivity analysis, along with any correlated parameters. Figure 4 shows the sensitivity analysis results of JTT source location for each M_w at each evaluation point. The source locations considered in this study exhibited significant variation in the north-south direction. Therefore, in Figure 4, the center point of fault y_{eq}, which roughly corresponds to the north-south position, is shown as converted latitude. At P1 and P2, where the tsunami propagating directly from the source produces the maximum tsunami height, the peak maximum tsunami height occurs when the source is approximately in front of the evaluation point. Note that at P2, the tsunami height shows a discontinuity near y_{eq} of 37°. This is considered to be the boundary between the case where the fault model spans the northern and southern sections of the Japan Trench (refer to Chapter 3) and the case where the fault model is a single fault in the southern section. However, even at this boundary, the tsunami height should ideally be continuous. Improving the modeling of this boundary area will be a subject of future research. At P3, the tsunamis propagating from the north and directly in front are obstructed by the Oshika Peninsula and arrive after diffracting around the peninsula. The maximum tsunami height is considered to be the complex superposition of multiple waves. On



(i) P1 (ii) P2

(iii) P3

Figure 4: JTT sensitivity analysis of source location (x_{eq}, y_{eq}) to tsunami height (θ= base, d= 1m)



(i) P1 (ii) P2

(iii) P3

Figure 5: JTT sensitivity analysis of M_w to tsunami height (Source location (x_{eq}, y_{eq}) fixed at front of each evaluation points, d= 1m)

Table 4: Standard deviation of tsunami height for varying each parameter (JTT)

Source	standard deviation σ				
	All parameter varied	Location (x_{eq}, y_{eq})	M_w	Strike θ	Fault depth d
P1	3.87	2.94	3.33	0.40	0.11
P2	5.01	3.72	3.53	0.28	0.21
P3	1.11	0.87	1.12	0.07	0.05

the other hand, the waves propagating from the south travel directly towards P3. When the source is located south of the P3 (latitude of y_{eq} is roughly south of 37.5°), the tsunami height tends to be higher compared to the case where it is located to the north.

Figure 5 shows the sensitivity analysis results for M_w at each evaluation point for JTT. At all evaluation points, higher M_w values resulted in greater tsunami heights. However, some cases exhibited non-linear trends, which are likely attributable to the interaction between the wave field and coastal topography. For example, in the case of θ = base + 10° at P2, the tsunami height is low when M_w is small because the tsunami propagates northward from the evaluation point. On the other hand, as M_w increases, the fault length increases in the north-south direction, then the wave field generated in the southern part of the fault propagates towards the evaluation point, resulting in a higher tsunami height.

Table 4 shows the standard deviation σ of the maximum tsunami height when only each parameter is varied. These analyses reveals that the source location and M_w have the largest contribution to the maximum tsunami height.

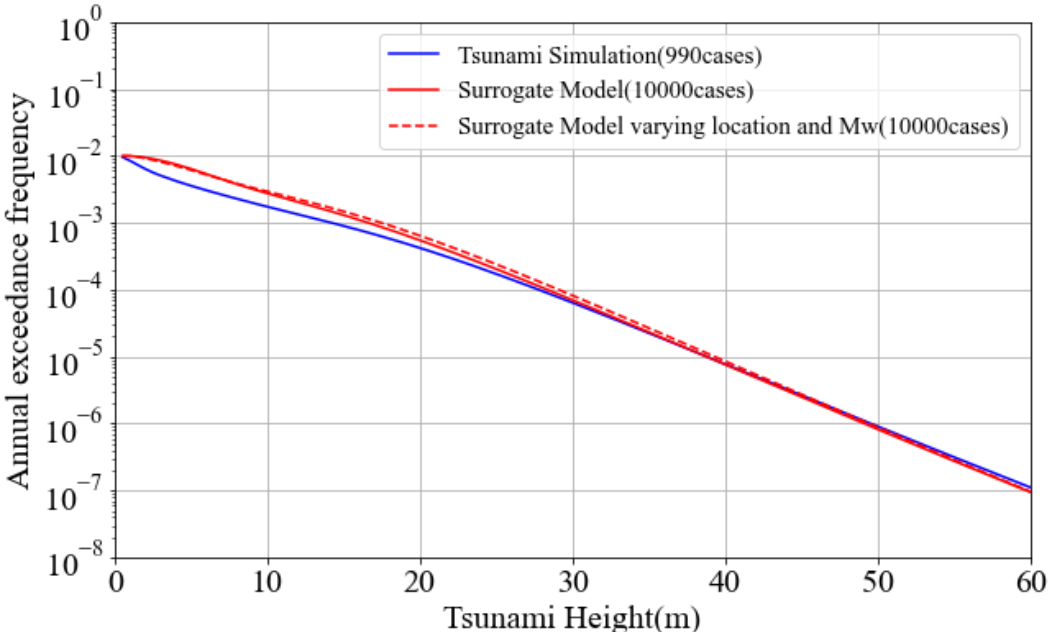


Figure 6: Annual exceedance frequency at P2 (Tsunami numerical calculation vs. Surrogate model)

6.3 Sensitivity analysis for probabilistic tsunami hazard curve

Probabilistic tsunami hazard curves for the maximum tsunami height at P2 for the JTT, obtained from both the surrogate model and tsunami numerical calculations, were generated based on the JSCE [9] method.

The fault parameters (source location, M_w, θ, d) were assumed to vary randomly according to continuous uniform distributions within the ranges shown in Table 1. Referring to JSCE [9], the mean recurrence interval was set to 150 years, the logarithmic normal distribution variability β was set to 0.30, and truncation was set to $\pm 10\beta$. While it would be possible to set the variability and truncation of the tsunami estimates separately for the surrogate model and the tsunami numerical calculation, we used a unified approach in this study. Note that epistemic uncertainty was not considered in this study.

Figure 6 shows the hazard curves obtained from both the surrogate model and the tsunami numerical calculation. The hazard curve from the surrogate model, which can capture the continuous variation of parameters, indicates higher hazard values for tsunami heights below 25 m compared to the numerical simulation. However, the hazard curves are nearly identical for tsunami heights above 25 m. Also shown in Figure 6 is the hazard curve obtained by varying only the source location and M_w, which have a significant contribution to the response uncertainty (refer to Table 4). θ was fixed at base°, and d at 1 m. At P2, the hazard curve for the case where only the source location and Mw were varied was almost identical to the case where all parameters were varied. This clearly indicates the significant contribution of source location and Mw to the hazard curve.

7 CONCLUSIONS

This study uses surrogate models trained on tsunami numerical calculation results at coastal line along the Pacific coast for tsunami sources along the Japan Trench. This model provides the probability distribution of the maximum tsunami height under continuous input parameters. Sensitivity analyses using the surrogate model revealed that the source location and M_w are the most significant contributors to the maximum tsunami height. A comparison of hazard curves from the surrogate model and tsunami numerical calculations was conducted. Future work will involve uncertainty assessments for not only the maximum tsunami height but also the period variation and tsunami direction. Furthermore, we plan to conduct an uncertainty assessment of the response due to features characterizing coastal topography.

REFERENCES

[1] Kotani, T., Tozato, K., Takase, S., Moriguchi, S., Terada, K., Fukutani, Y., Otake, Y., Nojima, K., Sakuraba, M. and Choe, Y.: Probabilistic tsunami hazard assessment with simulation-based response surfaces, *Coastal Engineering.* (2020) **160**:103719,

[2] Gopinathan. D., Heidarzadeh, M., and Guillas, S.: Probabilistic quantification of tsunami current hazard using statistical emulation, *Proceedings of the Royal Society A: Mathematical*, (2021) **477**:20210 180, https://doi.org/10.1098/rspa.2021.0180.

[3] Snelling, B., Neethling, S., Horsburgh, K., Collins, G. and Piggott, M.: Uncertainty Quantification of Landslide Generated Waves Using Gaussian Process Emulation and Variance-Based Sensitivity Analysis, *Water 2020*. (2020) **12**:416.

[4] Salmanidou, D. M., Beck, J., Pazak, P. and Guillas, S.: Probabilistic, high-resolution

tsunami predictions in northern Cascadia by exploiting sequential design for efficient emulation, *Nat. Hazards Earth Syst. Sci.* (2021) **21**:3789–3807.
[5] Williamson, A L., Rim, D., Adams, L M., LeVeque, R J., Melgar, D. and González, F I.: A Source Clustering Approach for Efficient Inundation Modeling and Regional Scale Probabilistic Tsunami Hazard Assessment, *Frontiers in Earth Science*. (2020) **8**:591663.
[6] Fukutani, Y., Yasuda, T. and Yamanaka, R.: Efficient probabilistic prediction of tsunami inundation considering random tsunami sources and the failure probability of seawalls, *Stochastic Environmental Research and Risk Assessment.* (2023) **37**:2053–2068.
[7] Kihara, N., Kaida, H., Takahashi, Y., Nishi, A., Kimura, T., Fujii, N., Fujioka, B., Oda, S., Ohtori, Y. and Mihara, Y.: Tsunami flooding analysis graded-approach framework for tsunami probabilistic risk assessment, *Journal of Nuclear Science and Technology.* (2023).
[8] Nishi, A., Kihara, N., Kimura, T., Fujii, N. and Shoji, G.: Modeling of tsunami scenarios for refinement of fragility assessment for nuclear power plants*, Transaction of the Japan Society of Civil Engineers.* (2024) **80**:13.
[9] Japan Society of Civil Engineers (JSCE): *Tsunami Assessment Method for Nuclear Power Plants in Japan* (2016).
[10] National Research Institute for Earth Science and Disaster Prevention, Japan (NIED): *An Approach to Tsunami Hazard Assessment along the Northeastern Coastal Area in Japan –Method and Preliminary Results-*. (2015).
[11] Nishi, A. and Shoji, G.: Influence assessments of various features for tsunami source and fault deformation distribution on tsunami height at shoreline ~case study at nuclear power plant into earthquakes along the Japan Trench~*, Transaction of the Japan Society of Civil Engineers.* (2025) **81**:13.
[12] Vapnik, V.: The nature of Statistical Learning Theory, *Springer.* (1995).
[13] Mochihashi, D., Oba, S.: Gaussian Process and Machine Learning, *Kodansha*. (2019).
[14] GPy. GPy: A Gaussian process framework in python since 2012, http://github.com/SheffieldML/GPy